# HPC Trends, Big Data And The Emerging Market For High Performance Data Analysis (HPDA)

Salishan April 2013

Earl Joseph, IDC
ejoseph@idc.com
www.hpcuserforum.com

# IDC Has >1,000 Analysts In 52 Countries

IDC
Analyze the Future

# Presentation Outline

- **IDC HPC Research Activities**
- **HPC Market Update and Trends**
  - HPC Usage by Major Countries
- **Major IT Trends**
- **IDC's Perspective On Big Data Outside Of HPC**
- **Big Data Meets HPC And Advanced Simulation**
  - End-Use Examples Of Big Data Today
- **New HPC, Innovation and ROI Research**

# About IDC:
# IDC HPC Activities

- **Track all HPC servers sold each quarter**
- **4 HPC User Forum meetings each year**
- **Publish 45 plus research reports each year**
- **Visit all major supercomputer sites & write reports**
- **Assist in collaborations between buyers/users and vendors**
- **Assist governments in HPC plans, strategies and direction**
- **Assist buyers/users in planning and procurements**
- **Maintain 5 year forecasts in many areas/topics**
- **Conduct special research studies**

# HPC Market Update And Trends

# IDC's Definition Of Technical Computing (aka HPC)

**IDC uses these terms to cover all technical servers used by scientists, engineers, financial analysts and others:**

- HPC
- HPTC
- Technical computing servers
- Highly computational servers

**HPC covers all servers that are used for highly computational or data intensive tasks**

# Top Trends in HPC

IDC Analyze the Future

**The global economy in HPC is growing again:**

- 2010 grew by 10%, to reach $9.5 billion
- 2011 grew by 8.4% to reach $10.3 billion
- HPC revenue for 2012 exceeded $11B
  - **Q3 2012 was the largest quarter ever in HPC**
- We are forecasting ~7% growth over the next 5 years

**Major challenges for datacenters**

- Power, cooling, real estate, system management
- Storage and data management continue to grow in importance

**Software hurdles continue to grow**

**The worldwide Petascale Race is in full speed**

**Big Data and accelerators are hot new technologies**

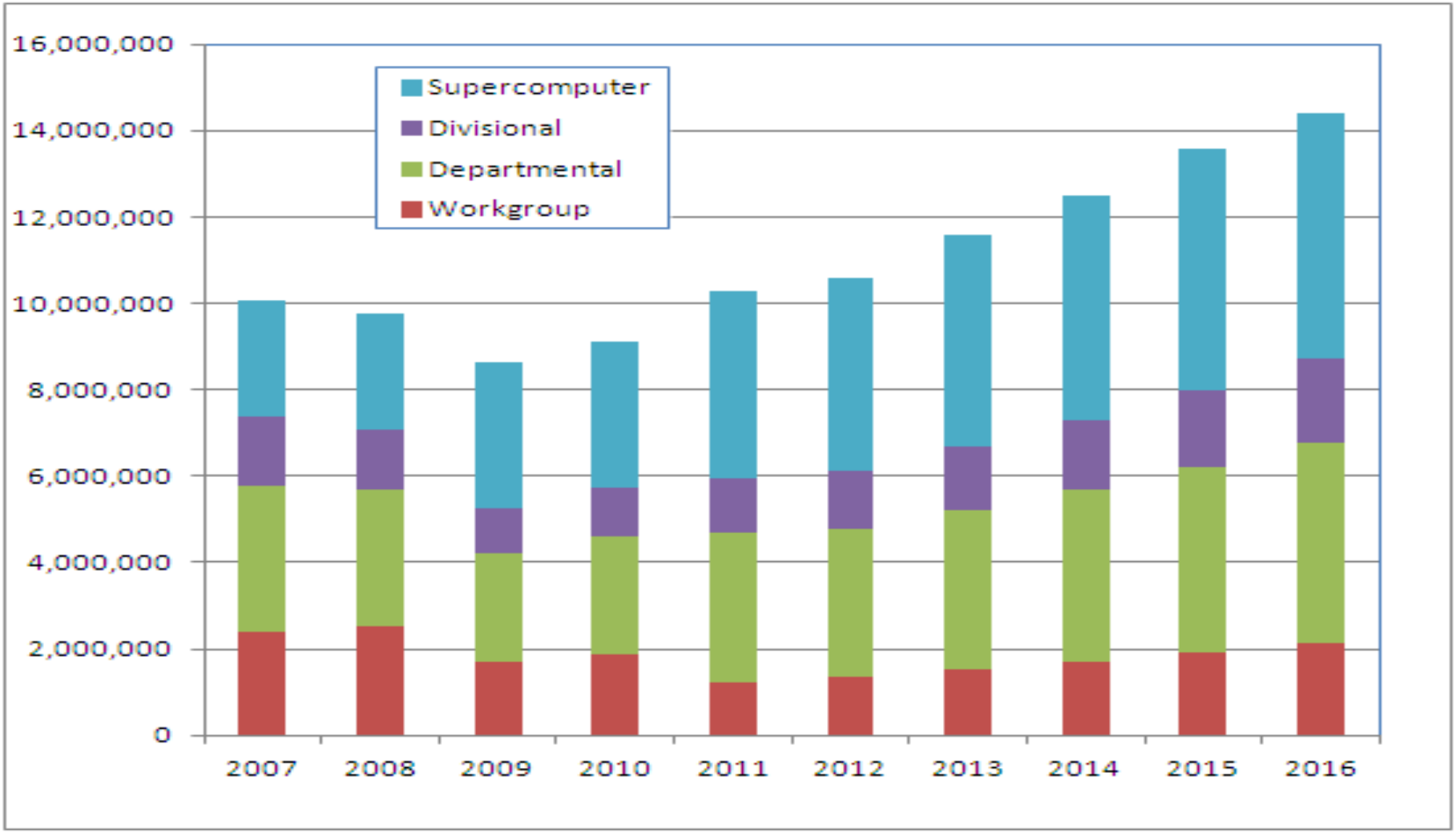
# HPC WW Market Trends: By Competitive Segments

| | 2008 | 2009 | 2010 | 2011 | CAGR '10/'11 |
|---|---|---|---|---|---|
| Supercomputer | 2,686,128 | 3,342,073 | 3,475,577 | 4,361,336 | 25.5% |
| Divisional | 1,395,817 | 1,078,575 | 1,268,735 | 1,245,541 | -1.8% |
| Departmental | 3,167,096 | 2,783,518 | 3,279,219 | 3,480,676 | 6.1% |
| Workgroup | 2,522,809 | 1,409,979 | 1,474,792 | 1,212,505 | -17.8% |
| Grand Total | 9,771,849 | 8,614,145 | 9,498,323 | 10,300,058 | 8.4% |

# HPC WW Market Trends: By System Units Sold

| | 2008 | 2009 | 2010 | 2011 | CAGR '10/'11 |
|---|---|---|---|---|---|
| Supercomputer | 1,863 | 2,067 | 2,560 | 2,893 | 13.0% |
| Divisional | 4,054 | 3,596 | 3,914 | 3,739 | -4.5% |
| Departmental | 20,105 | 17,098 | 19,868 | 20,770 | 4.5% |
| Workgroup | 148,069 | 82,293 | 93,502 | 84,149 | -10.0% |
| Grand Total | 174,091 | 105,054 | 119,844 | 111,551 | -6.9% |

# HPC Forecasts: By Competitive Segment



# The HPC Market Beyond The Servers: The Broader HPC Market

IDC Analyze the Future

| The Broader HPC Market Growth to 2016 | | | | | | | |
|---|---|---|---|---|---|---|---|
| HPC Compute, Storage, Middleware, Application and Service Revenues, 2011 -- 2016 ($M | | | | | | | |
| | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | CAGR (11-16) |
| Server | 10,300 | 11,031 | 11,910 | 12,778 | 13,839 | 14,621 | 7.3% |
| Storage | 3,664 | 3,992 | 4,350 | 4,739 | 5,163 | 5,625 | 8.9% |
| Middleware | 1,147 | 1,233 | 1,326 | 1,426 | 1,534 | 1,650 | 7.5% |
| Applications | 3,370 | 3,618 | 3,884 | 4,169 | 4,475 | 4,804 | 7.3% |
| Service | 1,801 | 1,924 | 2,056 | 2,197 | 2,348 | 2,509 | 6.9% |
| Total | 20,282 | 21,799 | 23,526 | 25,310 | 27,359 | 29,209 | 7.6% |
| Source: IDC 2012 | | | | | | | |

# National Competition For Supercomputing Leadership

# Why Is HPC Becoming So Important To Nations?

**High performance computing (HPC) is important for national economies, because HPC, also called supercomputing, has been firmly linked to economic competitiveness as well as scientific advances**

- **In one worldwide IDC study, 97% of companies that had adopted supercomputing said they could no longer compete or survive without it**
- **As the US COC puts it: To out-compute is to out-compete**

**Worldwide political leaders increasingly recognize this trend:**

- **In his 2006, State of the Union address, U.S. President George W. Bush promised to trim the federal budget, yet urged more money for supercomputing**

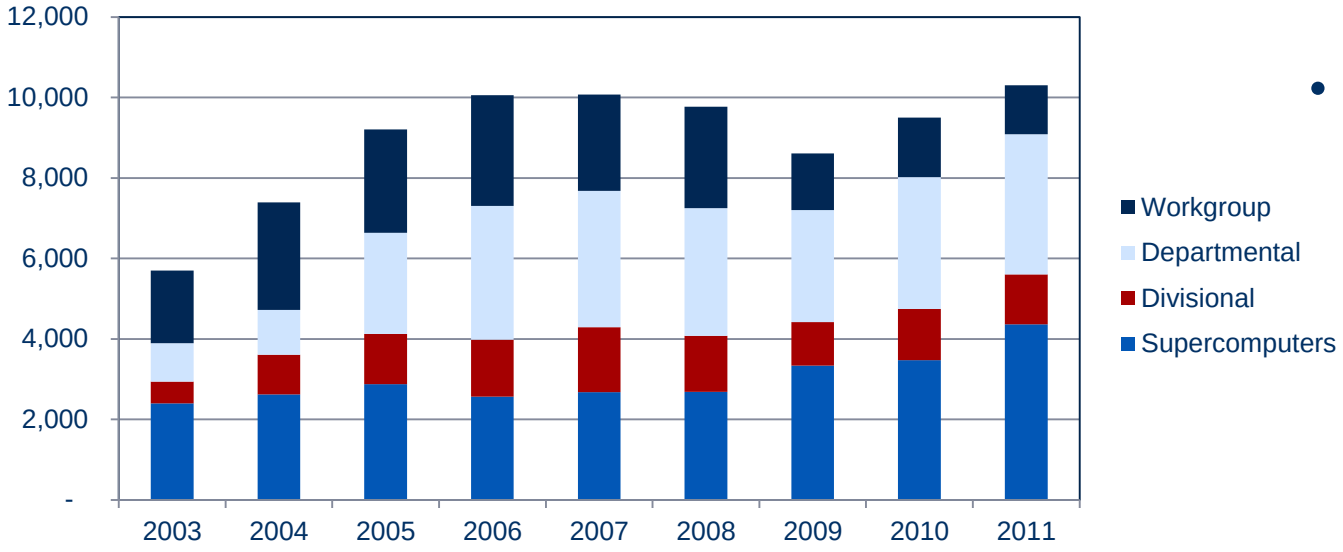
- **In 2009, Russian President Dmitry Medvedev warned that without more investment in supercomputer technology, "Russian products will not be competitive or of interest to potential buyers."**

# Why Is HPC Becoming So Important To Nations? (Continued)

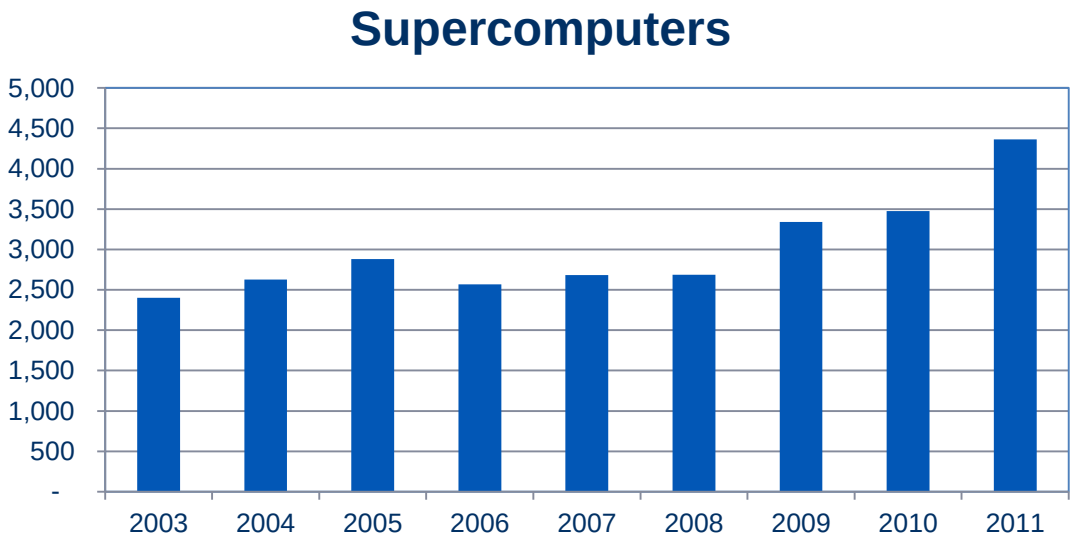
**Worldwide political leaders increasingly recognize this trend:**

- **In June 2010, Rep. Chung Doo-un of South Korea's Grand National Party: "If Korea is to survive in this increasingly competitive world, it must not neglect nurturing the supercomputer industry, which has emerged as a new growth driver in advanced countries."**
- **In his 2011 State of the Union address, President Obama noted China's rapid progress in HPC**
- **In February 2012, the European Commission announced that it has adopted a plan to double spending on HPC to €1.2 billion, with much of that money aimed at the installation of additional petascale supercomputer systems**
- **In January 2013, the EU launched two 1+ Billion euro, 10-year research investments (in brain research and graphene)**

# Supercomputer WW Growth Is Reshaping The Market

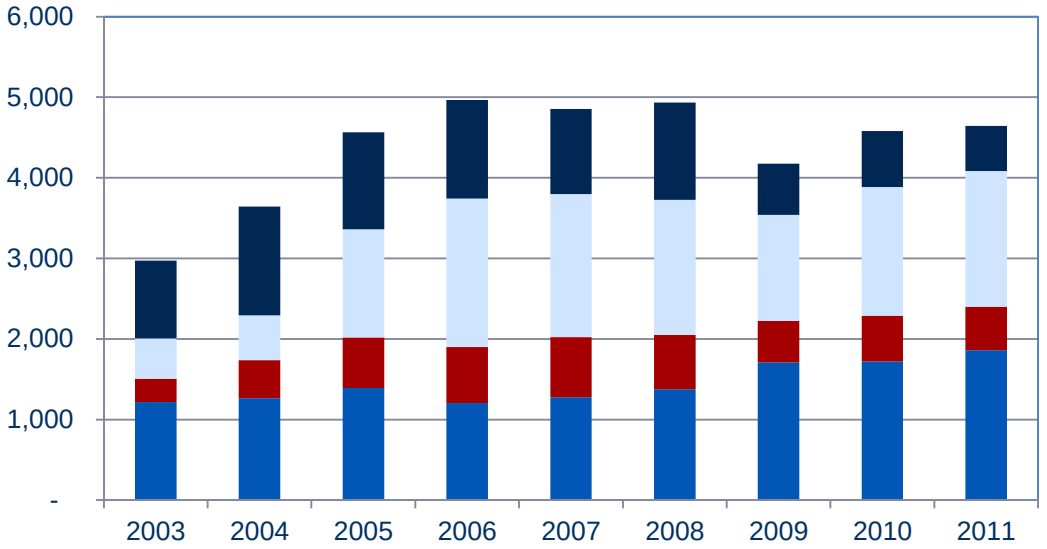


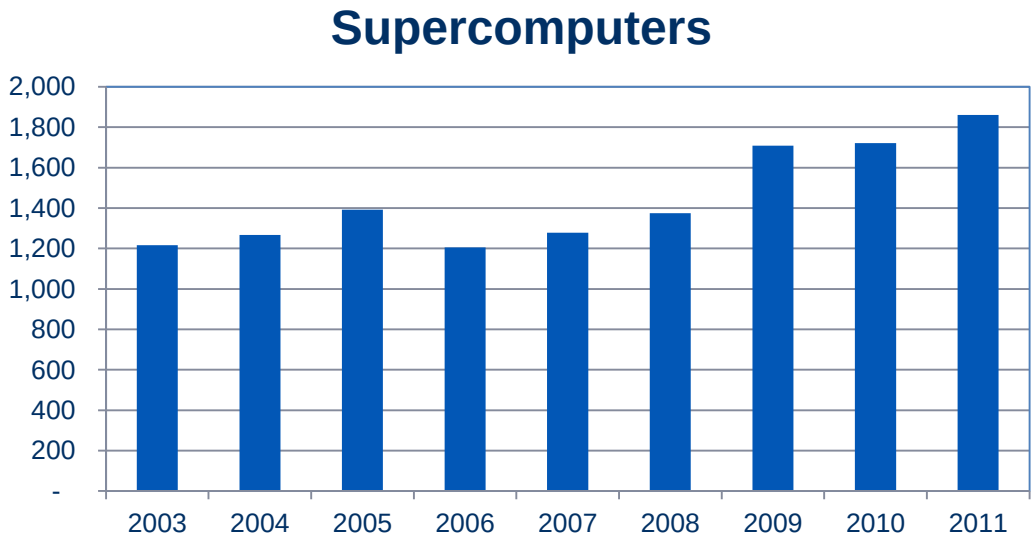

- **The overall HPC market was hard it by the recession, and has now fully recovered**




- **The worldwide supercomputer segment went into a major growth cycle – from 2008 to 2012**

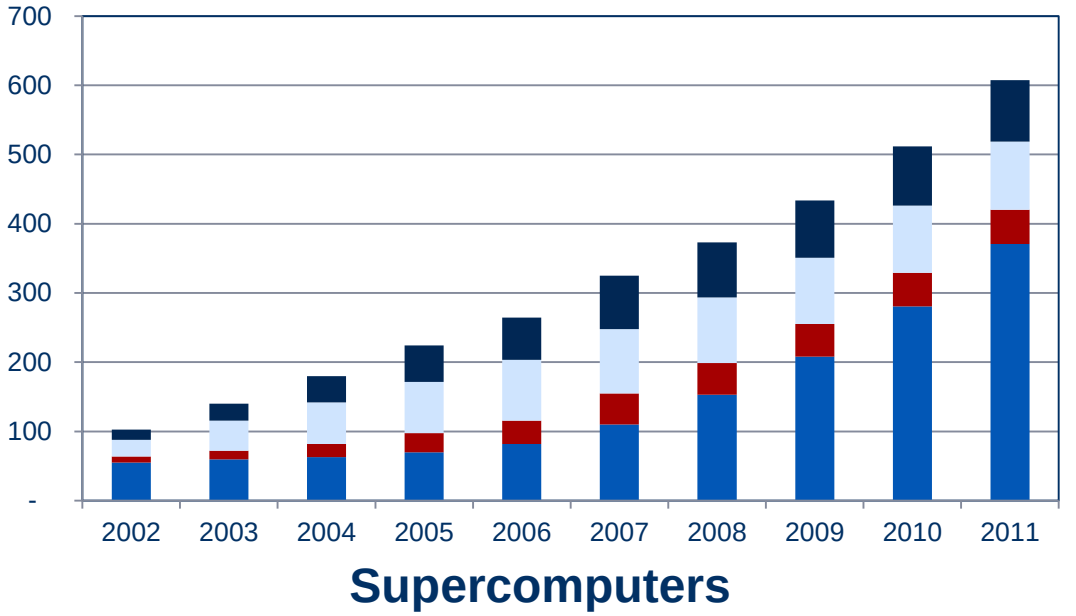
# USA Supercomputer Growth



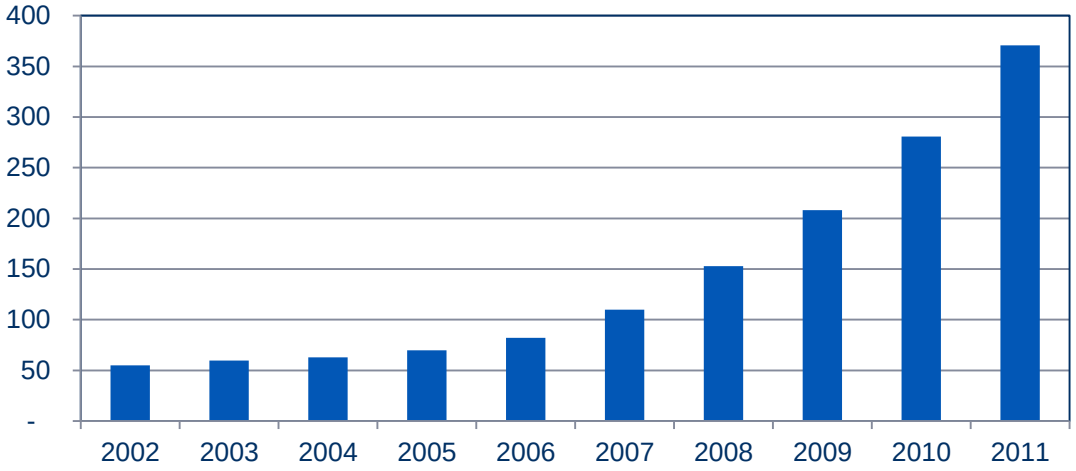




- **The overall USA HPC market was hard it by the recession, and still hasn't fully recovered**
- **The USA supercomputer segment grew well**

# China Supercomputer Growth

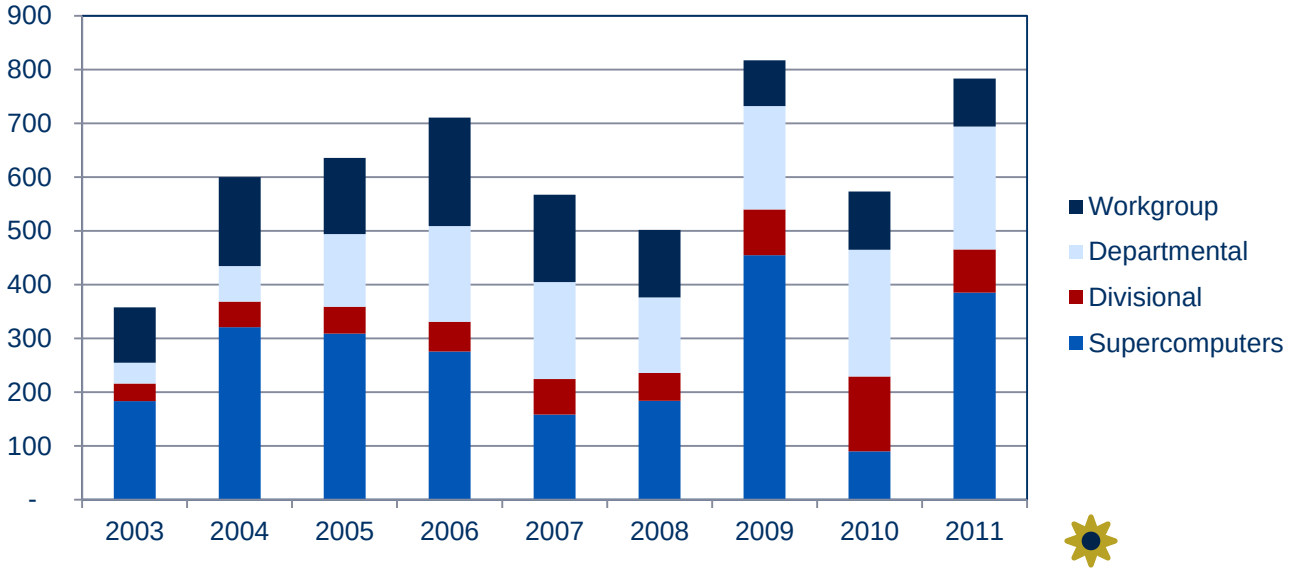


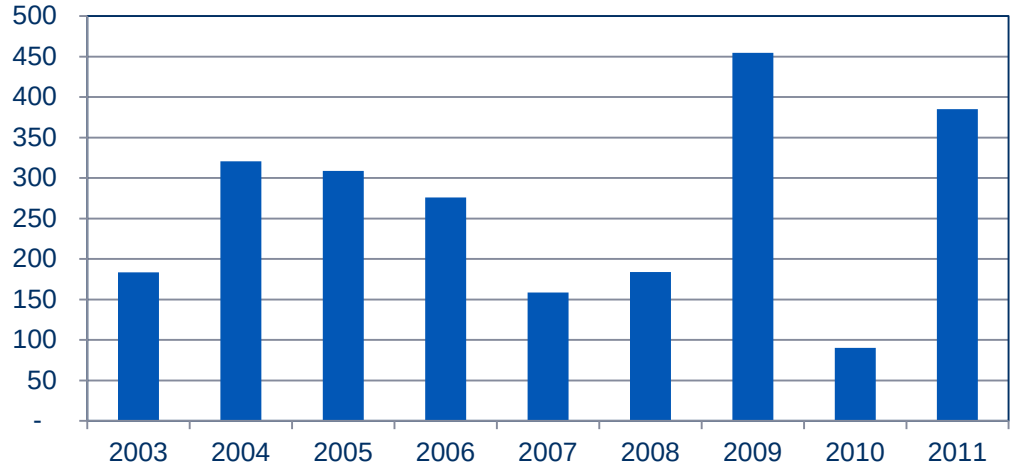

**Supercomputers**




- **The China HPC market wasn't impacted by the recession, and is well underway to reach $1 billion in 5 years**
- **The china supercomputer segment grew the most, heavily since 2007**

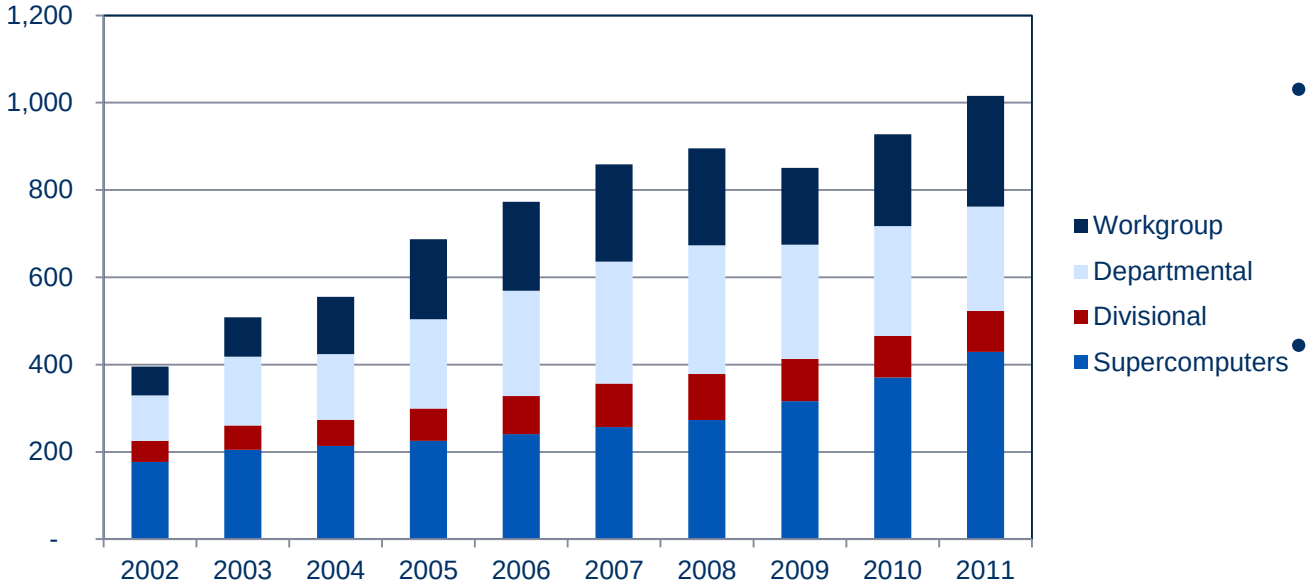
# Japan Supercomputer Growth



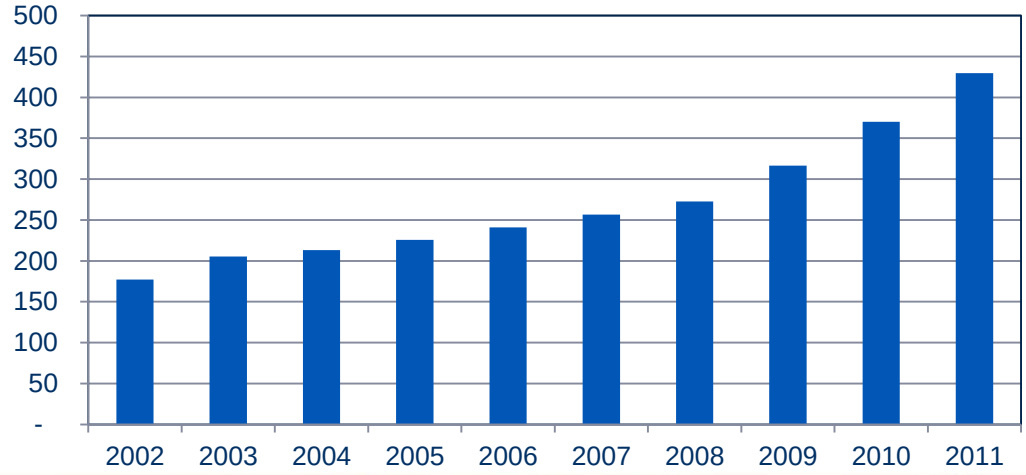

- **The Japan HPC market was hard it by the recession**
- **The Japan supercomputer segment is very up and down (Riken will make 2012 very large)**

# German Supercomputer Growth



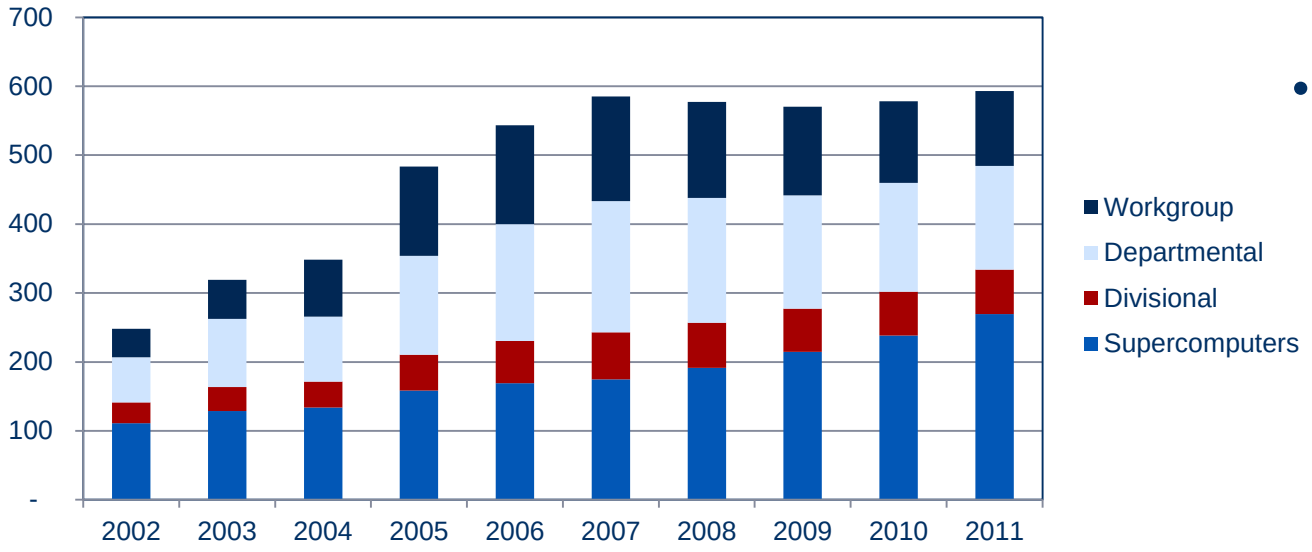

- **The German HPC market wasn't hit as much by the recession**
- **It just reached $1 billion in 2011**

**Supercomputers**

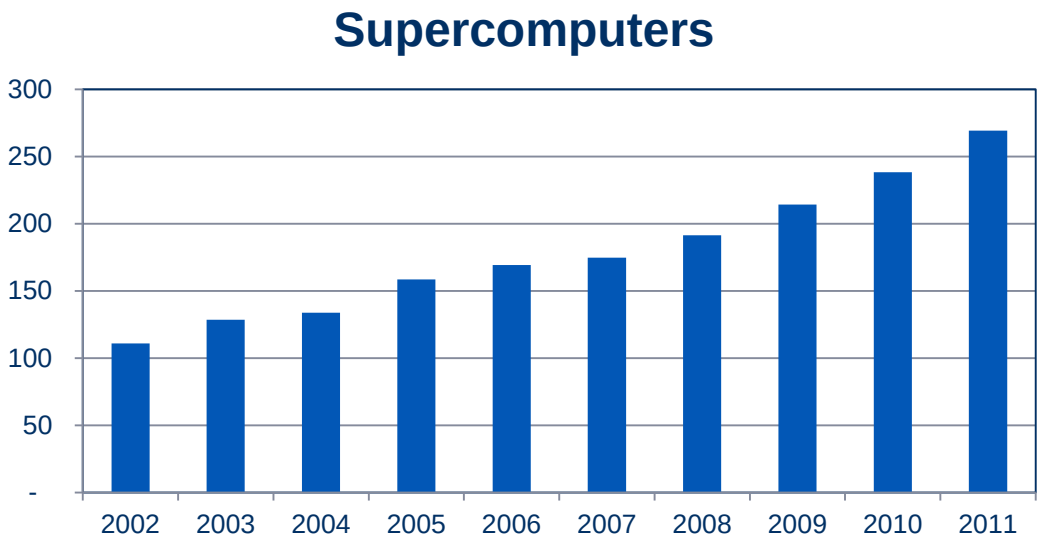



- **The German supercomputer segment is growing well**

# French Supercomputer Growth




- **The French HPC market was less impacted by the recession, and has been flat for 5 years**

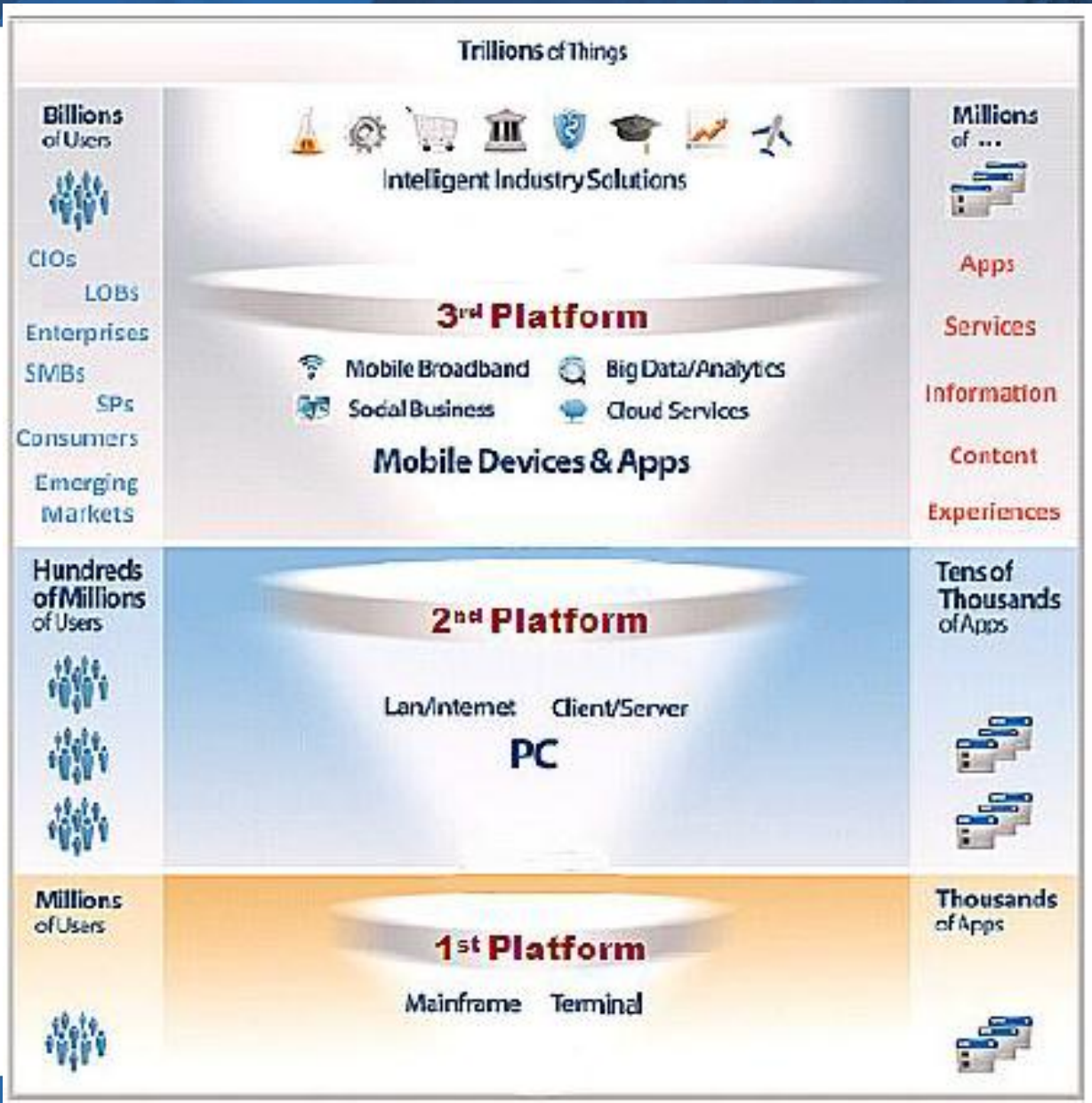



- **The French supercomputer segment is growing well**

# Major IT Trends

## *The Third Platform:*
## *The Intersection Of The Four Pillars Of Mobility, Social, Cloud And Big Data*

# IDC's View Of The Future Of IT: The Third Platform



# The Four Pillars

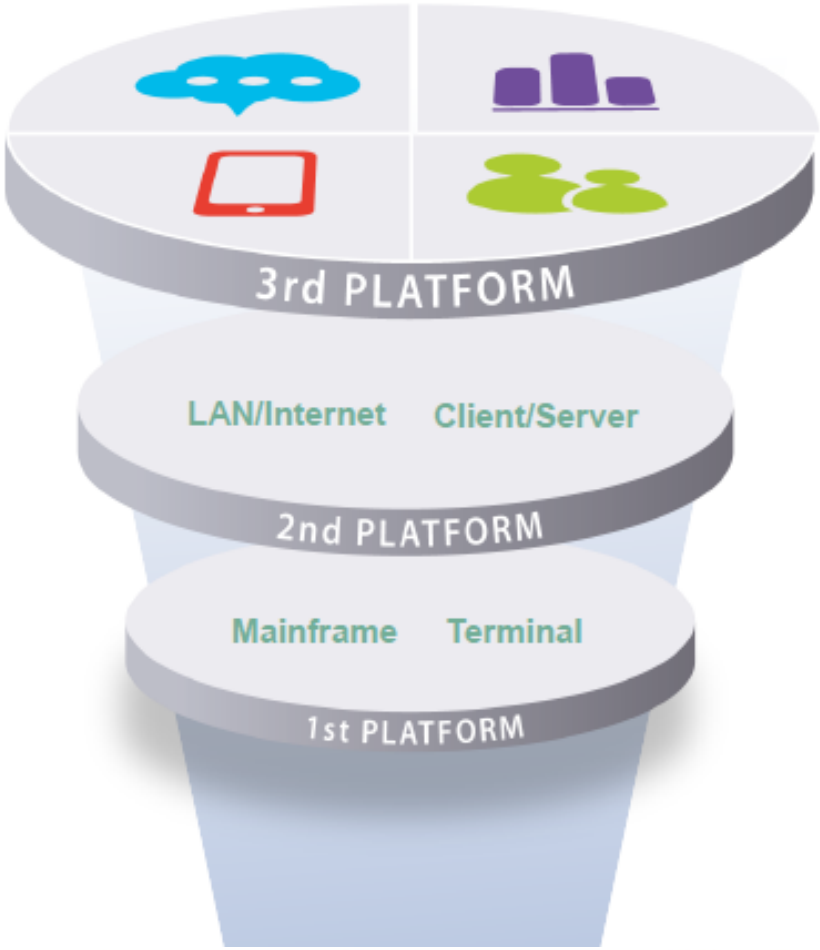
**IDC Four Pillar research provides insight into the intersection and impact of emerging intelligent solutions**

- As Big Data/Analytics, Cloud, Mobility and Social come together in an unprecedented way to create entirely new business value solutions and change the ICT playing field, they are enabling intelligent industries and solutions, and most of all - innovation.
- While productivity issues remain important, there is a noticeable shift across all industries in the decision making and purchasing when it comes to these new technologies.

# IDC's View Of The Future Of IT: The Third Platform

## The Growing Significance of "The 3rd Platform"

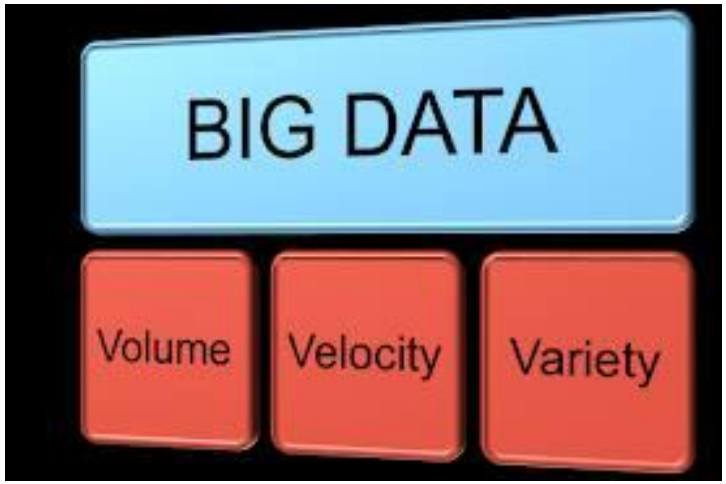
- Citizen/consumer transformation pushes enterprise transformation
- New platforms
- New engagement models
- New decision makers
- Continuing focus on operational efficiency
- New focus on
  - Outcomes versus outputs
  - Innovation




*Source: IDC 2012*

# IDC's Perspective On Big Data Outside Of HPC

# Big Data:
# A General Definition




+

**Value**

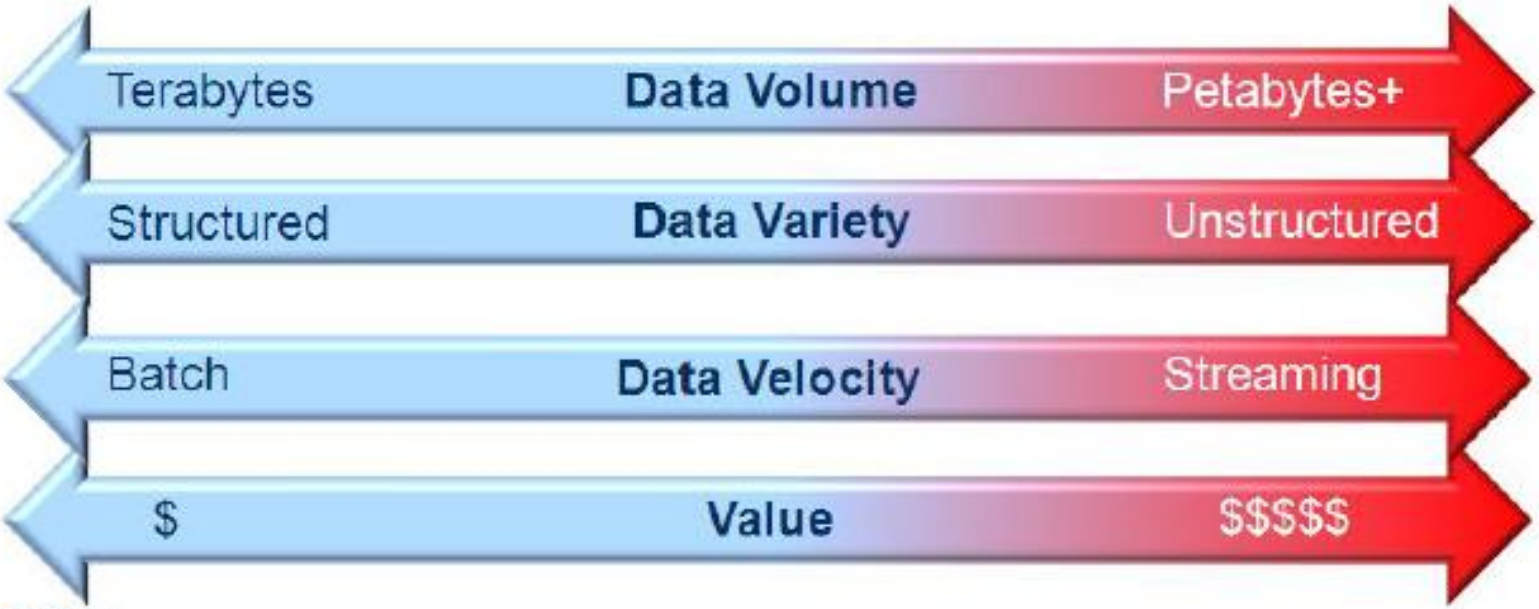
- Lots of data
- Time critical
- Multiple types (e.g., numbers, text, video)

- Worth something to someone

# Defining Big Data: For the Broader IT Market

IDC Analyze the Future

Big Data technologies describe a **new** generation of technologies and architectures, designed to **economically** extract **value** from very large **volumes** of a wide **variety** of data, by enabling high **velocity** capture, discovery and/or analysis.




*Source: IDC 2012*

# Top Drivers For Implementing Big Data

IDC Analyze the Future

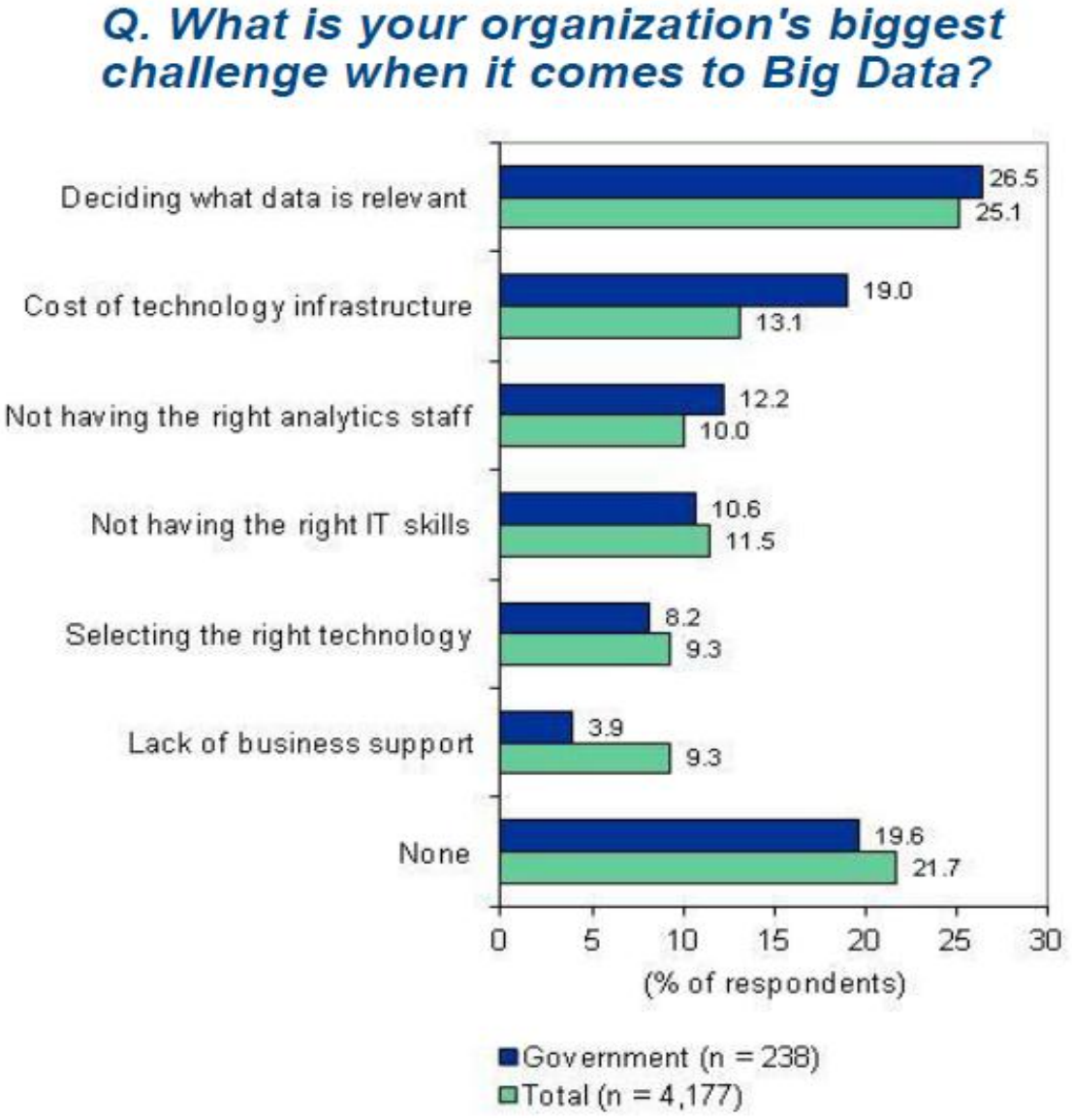
- 51.9% of government responders listed risk management, followed by cost control and workforce optimization
- Innovation and customer service trail as drivers

**Q. What were the top 3 drivers for your organization to implement Business Intelligence/Analytics solutions?**


*Source: 2012 IDC Vertical IT and Communications Survey*

# Organizational Challenges With Big Data: Government Compared To All Others

- Deciding what data is relevant was the top choice, followed by technology cost
- Skills shortage also shows up as a government concern



# Big Data Software

## Ecosystem Overview

### Big Data Software

**MPP RDBMS**

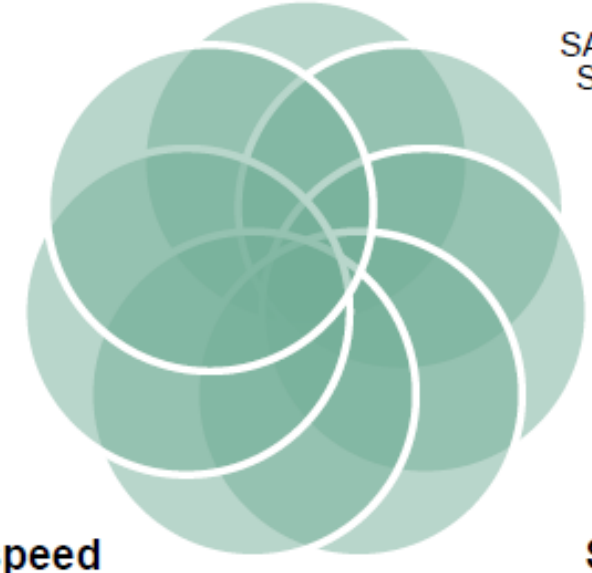
Oracle Exadata, IBM,Netezza Teradata, Microsoft, SAP, Sybase, EMC Greenplum , HP Vertica, ParAccel, InfoBright

**OODBMS**

Jade Software, Objectivity, Progress Software, Versant

**Predictive Analytics and Visualization**

SAS, IBM, Oracle, TIBCO, SAP, QlikTech, Tableau, KXEN, Skytree, Revolution Analytics Fuzzy Logix



**Graph DBs**

Neo Technology, Objectivity, Franz Inc. Sones, Ravel

**Hadoop +**

**Management, analysis, integration, applications**

Hstreaming, MapR, Oracle, IBM, Platfora, Zettaset, DataStax, Karmashere, Datameer, Cloudera, HortonWorks, Hadapt

**Ultra high speed streaming**

Informatica, IBM, Progress, TIBCO, SAP, Information Builders

**Search & Discovery**

Attivio, Oracle, IBM, HP, Microsoft, Vivisimo, ZyLAB, Sinequa

*Source: IDC 2012*

# Big Data Software Technology Stack

IDC Analyze the Future

**Decision Support & Automation Interface**

Applications with functionality required to support collaboration, scenario evaluation, risk management, and decision capture and retention.

**Analytics & Discovery**

This layer includes software for ad-hoc discovery, and deep analytics and software that supports real-time analysis and automated, rules-based transactional decision making.

**Data Organization & Management**

Refers to software that processes and prepares all types of data for analysis. This layer extracts, cleanses, normalizes, tags, and integrates data.

**Infrastructure**

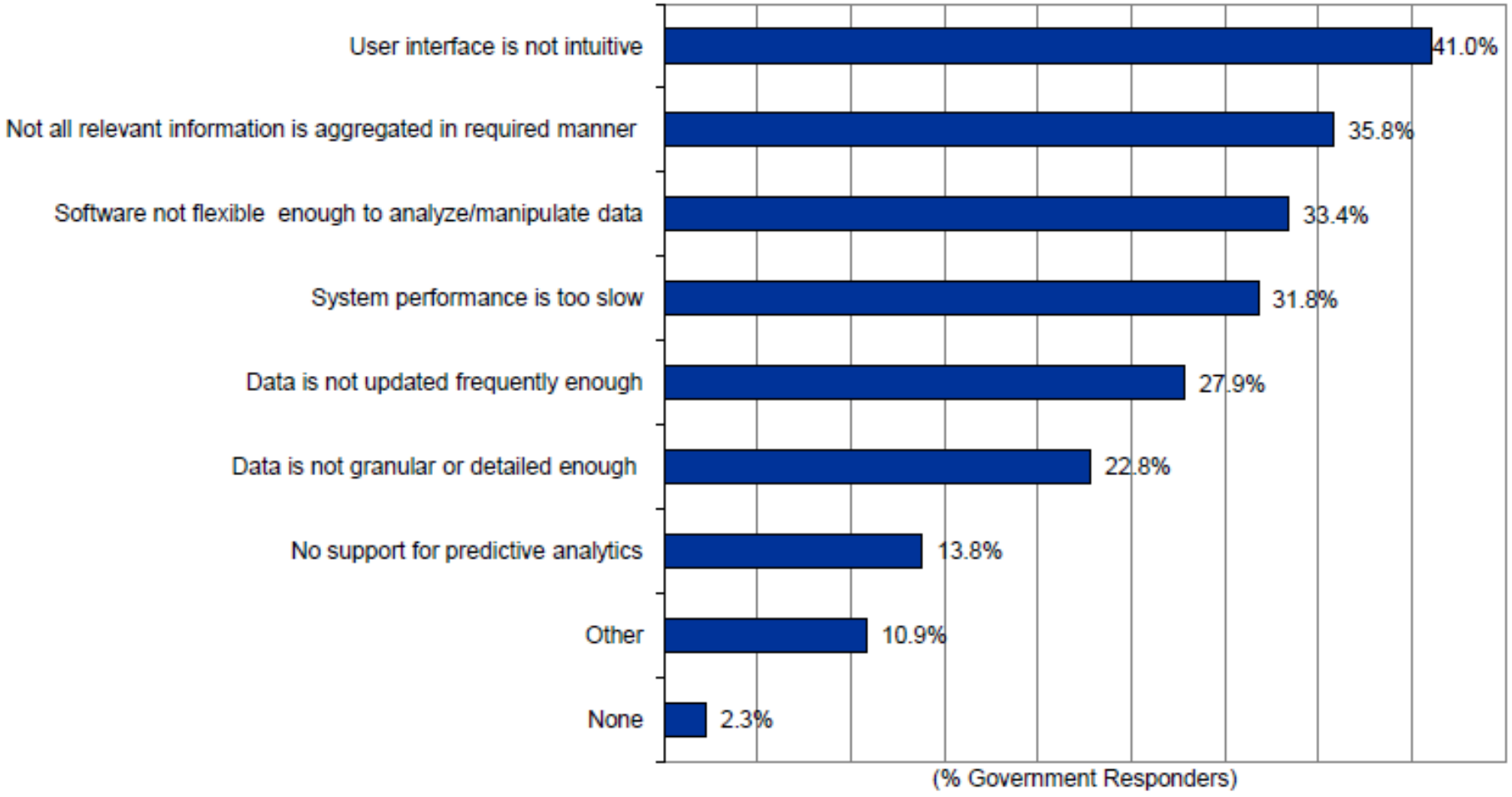
The foundation of the stack includes the use of industry-standard servers, networks, storage, and clustering software used for scale out deployment of Big Data technology.

*Source: IDC 2012*

# Big Data Software Shortcomings -- Today

IDC Analyze the Future

## *Q. What are top shortcomings of your Business Intelligence/Analytics solution(s)?*



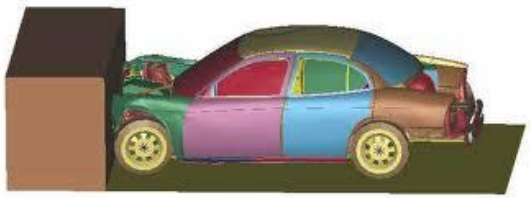
# BIG DATA MEETS HPC AND ADVANCED SIMULATION

# HPDA (High Performance Data Analysis): Data-Intensive Simulation and Analytics

IDC Analyze the Future

**HPDA = tasks involving sufficient data volumes and algorithmic complexity to require HPC resources/approaches**

- Established (simulation) or newer (analytics) methods
- Structured data, unstructured data, or both
- Regular (e.g., Hadoop) or irregular (e.g., graph) patterns
- Government, industry, or academia
- Upward extensions of commercial business problems
- Accumulated results of iterative problem-solving methods (e.g., stochastic modeling, parametric modeling).




# HPDA Market Drivers

- **More input data (ingestion)**
  - More powerful scientific instruments/sensor networks
  - More transactions/higher scrutiny (fraud, terrorism)
- **More output data for integration/analysis**
  - More powerful computers
  - More realism
  - More iterations in available time
- **The need to pose more intelligent questions**
  - Smarter mathematical models and algorithms
- **Real time, near-real time requirements**
  - Catch fraud before it hits credit cards
  - Catch terrorists before they strike
  - Diagnose patients before they leave the office
  - Provide insurance quotes before callers leave the phone

# Data Movement Is Expensive: In Energy and Time-to-Solution

IDC Analyze the Future

## Energy Consumption

- 1MW ≈ $1 million
- Computing 1 calculation ≈ 1 picojoule
- Moving 1 calculation = up to 100 picojoules
- => It can take 100 times more energy to move the results of a calculation than to perform the calculation in the first place.

## Strategies

- Accelerate data movement (bandwidth, latency)


- Minimize data movement (e.g., data reduction, in-memory compute, in-storage compute, etc.)

# Different Systems for Different Jobs

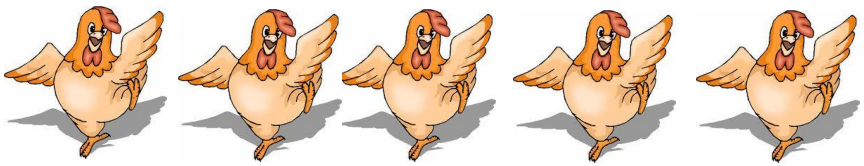
IDC Analyze the Future

## Partitionable Big Data Work

- Most jobs are here!
- Goal: **search**
- Regular access patterns (locality)
- Global memory not important
- Standard clusters + Hadoop, Cassandra, etc.

## Non-Partitionable Work

- Toughest jobs (e.g., graphing)
- Goal: **discovery**
- Irregular access patterns
- Global memory very important
- Systems turbo-charged for data movem



*versus*

HPC architectures today are compute-centric (FLOPS vs. IOPS)

# IDC HPDA Server Forecast

- Fast growth from a small starting point
- In 2015, conservatively approaching $1B

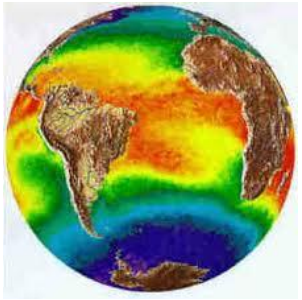
**TABLE 2**

IDC Worldwide Data Intensive (Big Data) Focused HPC Server Revenues

*($ Millions)*

| | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | CAGR '10-'15 |
|---|---|---|---|---|---|---|---|---|
| WW HPC Server Sales | 8,637 | 9,504 | 10,034 | 10,564 | 11,397 | 12,371 | 13,485 | 7.2% |
| Big Data Workloads | 535 | 603 | 655 | 708 | 786 | 881 | 989 | 10.4% |
| Big Data in HPC Portion | 6.2% | 6.3% | 6.5% | 6.7% | 6.9% | 7.1% | 7.3% | 3.0% |

Source: IDC 2012



# END-USE EXAMPLES OF BIG DATA TODAY

# Some Major Use Cases for HPDA

- **Fraud/error detection across massive databases**
  - A horizontal use – applicable in many domains
- **National security/crime-fighting**
  - SIGINT/anomaly detection/anti-hacking
  - Anti-terrorism (including evacuation planning)/anti-crime
- **Health care/medical informatics**
  - Drug design, personalized medicine
  - Outcomes-based diagnosis & treatment planning
  - Systems biology
- **Customer acquisition/retention**
- **Smart electrical grids**
- **Design of social network architectures**

# Use Case: PayPal Fraud Detection / Internet Commerce

**Slides and permission provided by PayPal, an eBay company**

# The Problem



Detecting fraud in 'real time' as millions of transactions are processed between disparate systems at volume.

Finding suspicious patterns that we don't even know exist in related data sets.

Ability to create and deploy new fraud models into event flows quickly and with minimal effort.

Provide environment for fraud modeling, analytics, visualization, M/R, dimensioning and further processing.

# What Kind of Volume?
# PayPal's Data Volumes And HPDA Requirements

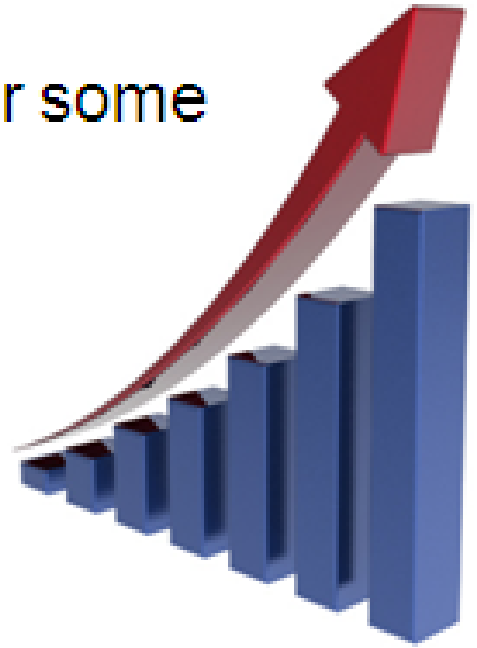
10 million+ logins / day

13 million financial transactions / day

300 variables calculated per event for some models

~4 Billion inserts / day

~8 Billion selects / day



# Where Paypal Used HPC

Infiniband on all internal Trinity network (Mellanox QDR 40Gb dual plane)

SGI InifiniteStorage IS4600 for EFL databases.

3 SGI Altix ICE 8200/8400 clusters for all 120+ EFL memory based apps – no disk i/o overhead.

MPI "like" apps. MPP features with scale out and affinity processing.

R&D with Lustre on Hadoop cluster and POC of columnar based database.

# The Results

- **$710 million saved in fraud that they wouldn't have been able to detect before (in the first year)**

# GEICO: Real-Time Insurance Quotes

- **Problem:** Need accurate automated phone quotes in 100ms. They couldn't do these calculations nearly fast enough on the fly.
- **Solution:** Each weekend, use a new HPC cluster to pre-calculate quotes for every American adult and household (60 hour run time)

# Something To Think About -- GEICO: Changing The Way One Approaches Solving a Problem

IDC Analyze the Future

- **Instead of processing each event one-at-a-time, process it for everyone on a regular basis**
  - It can be dramatically cheaper, faster and offers additional ways to be more accurate
  - But most of all it can create new and more powerful capabilities
- **Examples:**
  - Gun control background checks – calculate for every adult a simple yes or no – then send a coded database to all sheriffs each day
  - For home loan applications – calculate for every adult in the US and every home in the US
  - For health insurance fraud – track every procedure done on every US person by every doctor – and find patterns

# Something To Think About -- GEICO: Changing The Way One Approaches Solving a Problem

- **Examples (continued):**
- **In the future:**
  - If you add-in large scale data collection via sensors like GPS, drones and RFID tags:
    - New car insurance rules – The insurance company doesn't have to pay if you break the law -- like speeding and having an accident
    - You could track every car at all times – then charge $2 to see where the in-laws are in traffic if they are late for a wedding
    - Google maps could show in real-time where every letter and package is located
    - But crooks could also use it in many ways – e.g. watching ATM machines, looking for when guards are on break, …

# Global Courier Service: Fraud/Error Detection

**Here's a real-world example of one of the biggest names in global package delivery.**

**Their problem is not so different from PayPal's.**

**This courier service is doing real-time fraud detection on huge volumes of packages that come into their sorting facility from many locations and leave the facility for many other locations around the world.**

- **Check 1 billion-plus packages per hour in central sorting facility**
- **Benchmark won by a HPC vendor with a turbo-charged interconnect and memory system**

# Apollo Group/University of Phoenix: Student Recruitment and Retention

Apollo Group is approaching 300,000 online students. To maintain and grow, they have to target millions of prospective students.

Of course, it's no good to spend all that time and money on recruiting students if you can't keep most of them. Student losses, or "churn," are extremely expensive. So, the folks at Apollo also monitor enrolled students to spot signs of trouble and to intervene in some cases, such as suggesting a talk with a school counselor. The monitoring function relies on very sophisticated algorithmic models.

- Must target millions of potential students
- Must track student performance for early identification of potential dropouts – "churn" is very expensive
- Solution: a sophisticated, cluster-based Big Data models

SCHRÖDINGER

IDC
Analyze the Future


**Schrödinger is  a global life sciences software company.**

**One of the major things they do is use molecular dynamics to identify  promising candidates for new drugs to combat cancer and other diseases.**

**They use the cloud for this High Performance Data Analysis problem -- that's not so surprising, since molecular dynamics codes are often highly parallel.**

# Architecture


- They are using HPC in their on-premise data center, but the resources weren't big enough for this task.
- That's why they bursted out to Amazon EC2 using a software management layer from Cycle Computing to access more than 50,000 additional cores.
- Bringing a new drug to market can cost as much as £10 billion and a decade of time, so security is a major concern with commercial drug discovery.

# Optum + Mayo Initiative to Move Past Procedures-Based Healthcare

**You may have seen the recent news that Optum, which is part of United Health Group, is teaming with the Mayo Cline to build a large center ($500K) in Cambridge, Massachusetts to lay the research groundwork for outcomes-based medicine.**

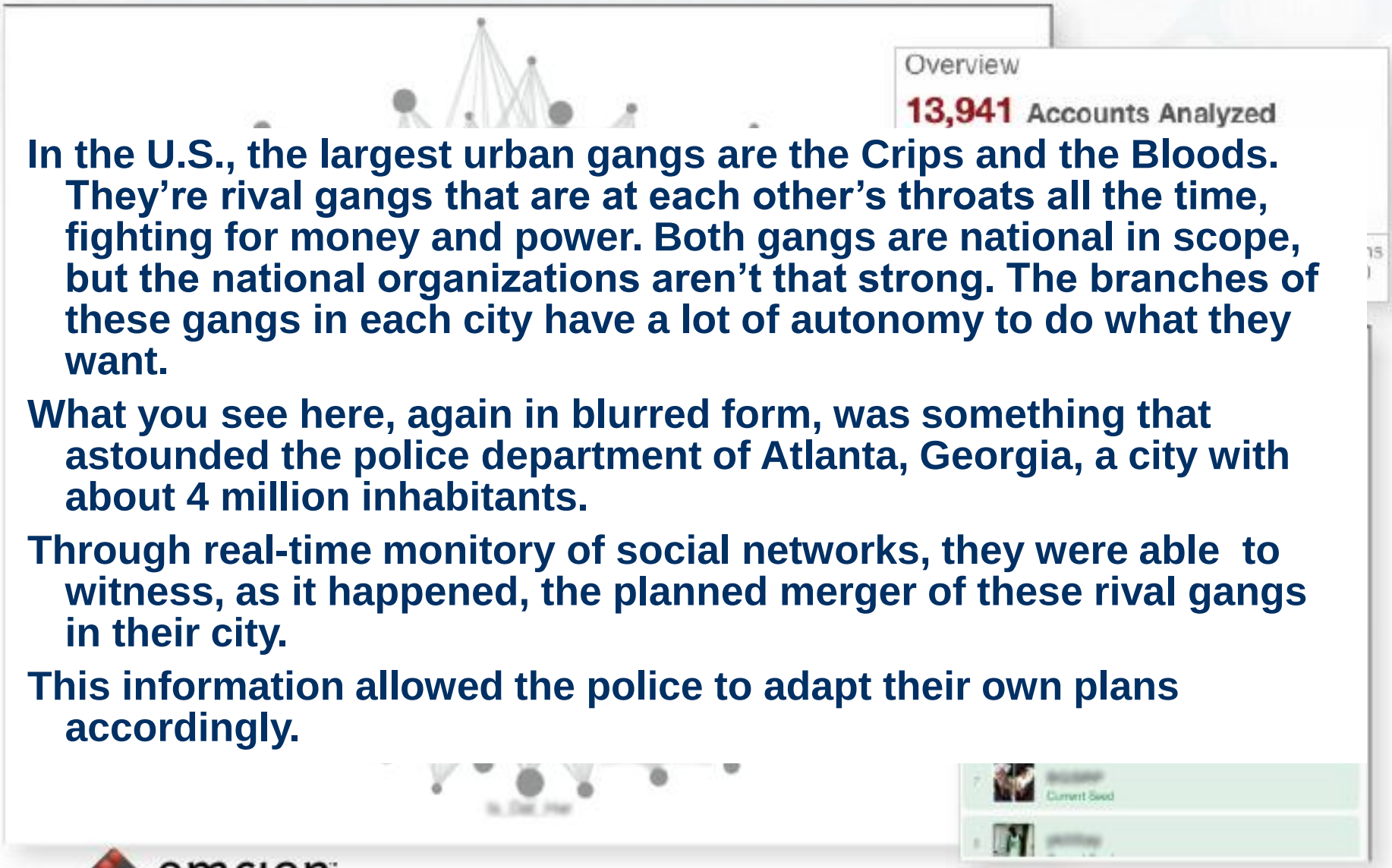
**They'll have more than 100 million patient records at their disposal for this enormous data-intensive work.**

**They'll be using data-intensive methods to look at other aspects of health care, too.**

**United Health issued a press release in which they said: "… they believe that improved efficiencies alone could reduce Medicare costs by about 40%, potentially obviating much of the need for the major reforms the political parties have been fighting about."**

- Data: 100M United Health Group claims (20 years) + 5M Mayo Clinic archived patient records. Option for genomic data
- Findings will be published
- Goal: outcomes-based care

# Use Case: Atlanta Gang Members Network Analysis




- In the U.S., the largest urban gangs are the Crips and the Bloods. They're rival gangs that are at each other's throats all the time, fighting for money and power. Both gangs are national in scope, but the national organizations aren't that strong. The branches of these gangs in each city have a lot of autonomy to do what they want.
- What you see here, again in blurred form, was something that astounded the police department of Atlanta, Georgia, a city with about 4 million inhabitants.
- Through real-time monitory of social networks, they were able to witness, as it happened, the planned merger of these rival gangs in their city.
- This information allowed the police to adapt their own plans accordingly.

emcien
Converting Data into Value


www.emcien.com | 7

# Summary: HPDA Market Opportunity

- **HPDA: simulation + newer high-performance analytics**
  - IDC predicts fast growth from a small starting point
- **HPC and high-end commercial analytics are converging**
  - Algorithmic complexity is the common denominator
  - Technologies will evolve greatly
- **Economically important use cases are emerging**
- **No single HPC solution is best for all problems**
  - Clusters with MR/Hadoop will handle most but not all work (e.g., graph analysis)
  - New technologies will be required in many areas
- **IDC believes our growth estimates could be conservative**

# HPDA User Talks: HPC User Forums, UK, Germany, France, China and U.S. … Let Us Know If You Want A Copy

- HPC in Evolutionary Biology, Andrew Meade, University of Reading
- HPC in Pharmaceutical Research: From Virtual Screening to All-Atom Simulations of Biomolecules, Jan Kriegl, Boehringer-Ingelheim
- European Exascale Software Initiative, Jean-Yves Berthou, Electricite de France
- Real-time Rendering in the Automotive Industry, Cornelia Denk, RTT-Munich
- Data Analysis and Visualization for the DoD HPCMP, Paul Adams, ERDC
- Why HPCs Hate Biologists, and What We're Doing About It, Titus Brown, Michigan State University
- Scalable Data Mining and Archiving in the Era of the Square Kilometre Array, the Square Kilometre Array Telescope Project, Chris Mattmann, NASA/JPL
- Big Data and Analytics in HPC: Leveraging HPC and Enterprise Architectures for Large Scale Inline Transactional Analytics in Fraud Detection at PayPal, Arno Kolster, PayPal, an eBay Company
- Big Data and Analytics Vendor Panel: How Vendors See Big Data Impacting the Markets and Their Products/Services, Panel Moderator: Chirag Dekate, IDC
- Data Analysis and Visualization of Very Large Data, David Pugmire, ORNL
- The Impact of HPC and Data-Centric Computing in Cancer Research, Jack Collins, National Cancer Institute
- Urban Analytics: Big Cities and Big Data, Paul Muzio, City University of New York
- Stampede: Intel MIC And Data-Intensive Computing, Jay Boisseau, Texas Advanced Computing Center
- Big Data Approaches at Convey, John Leidel
- Cray Technical Perspective On Data-Intensive Computing, Amar Shan
- Data-intensive Computing Research At PNNL, John Feo, Pacific Northwest National Laboratory
- Trends in High Performance Analytics, David Pope, SAS
- Processing Large Volumes of Experimental Data, Shane Canon, LBNL
- SGI Technical Perspective On Data-Intensive Computing, Eng Lim Goh, SGI
- Big Data and PLFS: A Checkpoint File System For Parallel Applications, John Bent, EMC
- HPC Data-intensive Computing Technologies, Scott Campbell, Platform/IBM
- The CEA-GENCI-Intel-UVSQ Exascale Computing Research Centre, Marie-Christine Sawley, Intel

# A New IDC Study:

# Creating Economic Models For HPC and ROI And for HPC And Innovation

# Background: Project Overview

**A study that describes how increases in HPC investments can significantly improve economic success and increase scientific innovation**

**The study includes creating two unique models:**

1. A macroeconomic model which depicts the way HPC investments result in economic advancements in the form of ROI, growth and jobs

2. An "Innovation Index" that provides a means of measuring and comparing innovation levels, based on the level of applying HPC computing resources towards scientific and technical advancement

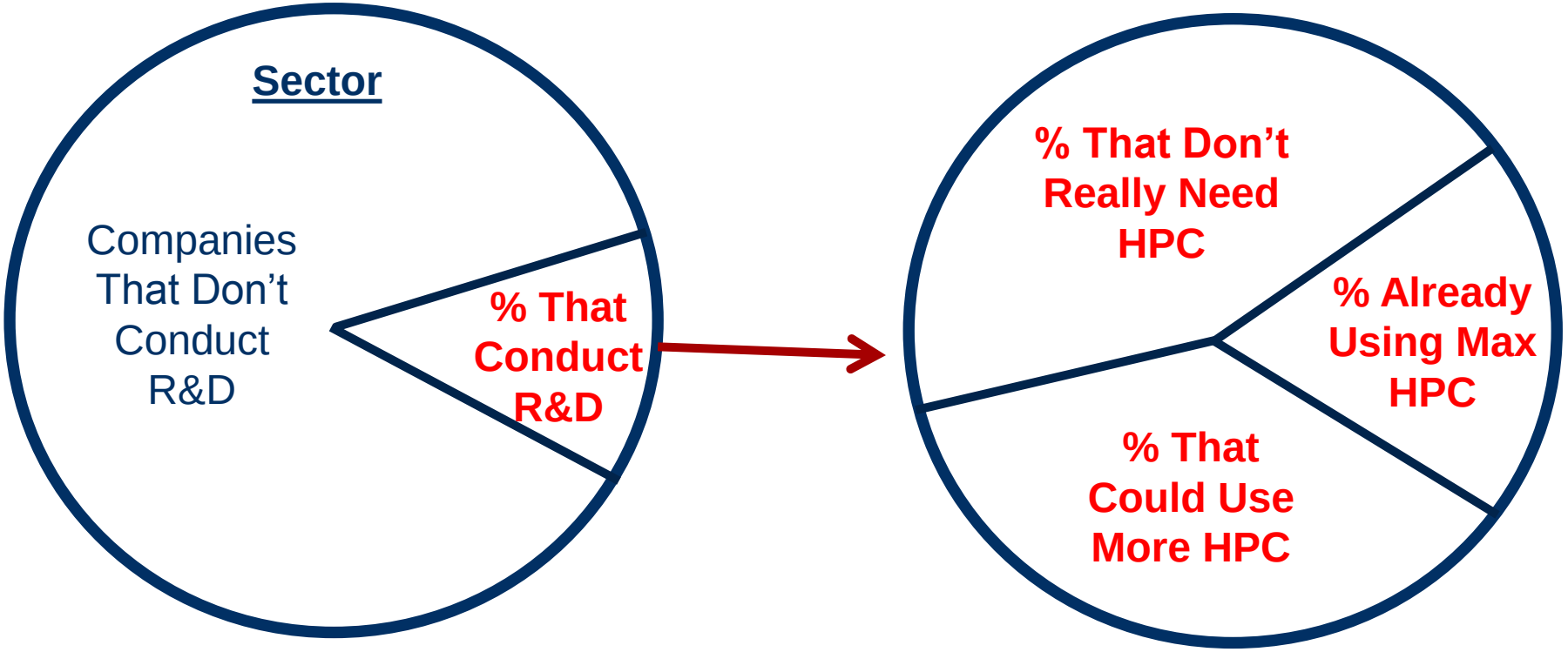
# Project Overview: Why It Is Key To DOE

IDC Analyze the Future

- World scientific leadership and innovation leadership are becoming more dependent on the use of HPC/supercomputers every year
- Economic leadership increasingly directly results from a nation's or an industry's or an enterprise's application of supercomputers in innovative and productive ways
- Other countries are putting into place plans to gain leadership in innovation and economic progress by more broadly applying HPC/supercomputing across many different industry and segments (like China, Russia, Europe and other Asian countries)

# Research Overview – An Example of the Parameters Needed

**For each sector we need 4 basic ratios**




Note: IDC has conducted over 10,000 light surveys for this data as of March 2013, it will likely require 5x to 10x more surveys

# The Initial Models That Are Being Tested

**The ROI models will be tested for two main types:**

1. **ROI based on revenues generated (similar to GDP) divided by HPC investment**
2. **ROI based on jobs created divided by HPC investment**

The ROI models will be tested for variances by:

1. Industry sector
2. Country
3. Organization size

**The Innovation models will be of four types:**

1. **Basic Research/Major Innovations in government & academia**
2. **Basic Research/Major Innovations in industry**
3. **Applied Research/Incremental Innovations in government & academia**
4. **Applied Research/Incremental Innovations in industry**

Two types of results will be used in the models:

1. Overall Importance of the innovation
2. And job creation

The Innovation models will be tested for variances by:

1. Industry sector
2. Country
3. Organization size
4. Government, Industry and Academia

# The Innovation Index Scale

**How would you rate this innovation compared to all other innovations in this field over the last ten years, using a scale of 1 to 10?**

**10 = One of the top 2 to 3 innovations in the last decade**

**9 = One of the top 5 innovations in the last decade**

**8 = One of the top 10 innovations in the last decade**

**7 = One of the top 25 innovations in the last decade**

**6 = One of the top 50 innovations in the last decade**

**5 = It had a major impact and is useful to many organizations**

**4 = A minor innovation that is useful to many organizations**

**3 = A minor innovation or only useful to 2 -3 organizations**

**2 = A minor innovation or only useful to 1 organization**

**1 = An innovation that is recognized ONLY by experts in the field**

# Research Schedule: 9/1/12 to June/July 2013

IDC Analyze the Future

1. ✓ **October 2012 – finalize the plan, surveys and overall research approach**
2. ✓ **November 2012 – conduct a limited number of surveys to both see what can be collected and start testing the two models**
3. ✓ **December 2012 – refine the survey guide and models as needed**
   - Refine as needed
4. ✓ **January – March 2013 – start the broad data collection/surveys – and review the initial results**
   - We noticed that we need a larger data set
5. Started **April/May 2013 – populate the two models with MORE DATA from both surveys and existing economic data elements**
   - Refine as needed
6. **May/June 2013 – present and refine the models as required, and create the 3 funding scenarios**
7. **June/July 2013 – disseminate the results**

# Questions?

Please email:
hpc@idc.com

Or check out:
www.hpcuserforum.com

# HPC Forecasts: By Verticals/Application Areas

IDC Analyze the Future

| | 2010 | 2011 | 2016 | CAGR (11-16) |
|---|---|---|---|---|
| Bio-Sciences | $1,240,127 | $1,251,665 | $1,722,588 | 6.6% |
| CAE | $1,013,233 | $1,095,398 | $1,714,457 | 9.4% |
| Chemical Engineering | $193,759 | $192,789 | $251,392 | 5.5% |
| DCC & Distribution | $519,549 | $569,026 | $868,925 | 8.8% |
| Economics/Financial | $253,607 | $279,294 | $472,015 | 11.1% |
| EDA / IT / ISV | $594,187 | $662,674 | $1,009,535 | 8.8% |
| Geosciences | $579,355 | $653,859 | $906,900 | 6.8% |
| Mech Design and Drafting | $75,316 | $63,102 | $79,128 | 4.6% |
| Defense | $919,558 | $1,004,632 | $1,380,750 | 6.6% |
| Government Lab | $1,467,110 | $2,078,029 | $2,714,603 | 5.5% |
| University/Academic | $1,762,777 | $1,900,883 | $2,526,773 | 5.9% |
| Weather | $388,735 | $453,999 | $601,585 | 5.8% |
| Other | $108,912 | $94,708 | $137,736 | 7.8% |
| Total Revenue | $9,116,225 | $10,300,058 | $14,386,387 | 6.9% |